

Speech Technology and the Wireless Internet

Mike Phillips
SpeechWorks International



The Evolution of Mobile Self-Service



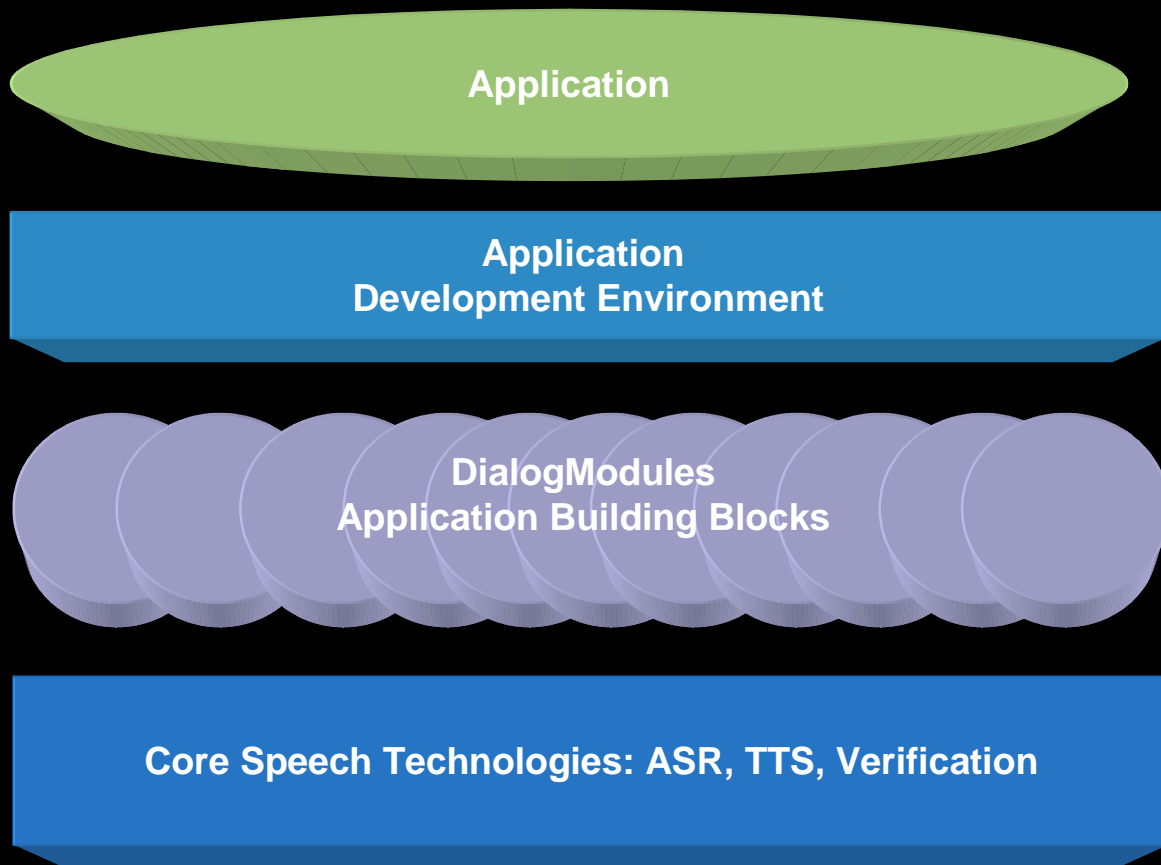
**Degree of Power
and Control**

Ease of Access and Use

Source: Gartner Group, 2000

SpeechWorks

Overall Architecture



State-of-the-art Speech Products

Continuous & speaker-independent

Phonetic -- type in words

Barge-in

Large vocabulary

Natural language through BNF grammars

Dynamic vocabularies and grammars

MultiLingual

Adapts to particular deployments

New Speech Technologies

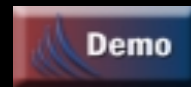


Extremely Large Vocabulary

- Vocabulary sizes >1 million words
 - Current dictionaries use 50 – 100K words
- Less computation and memory
 - Improves efficiency and accuracy
- Enables new generation of functionality
- Demo
 - Accessing directory-based information
 - Dialog technologies used to disambiguate answers from a database

“How May I Help You?”

- What is “How May I Help You?”?
 - Automated handling of highly unconstrained customer input via interactive dialogue
 - More flexible than today’s NL grammars
 - Leverages years of AT&T call center experience
- Enables a new class of speech applications
 - Call Routing
 - Help Desk
 - Customer Care

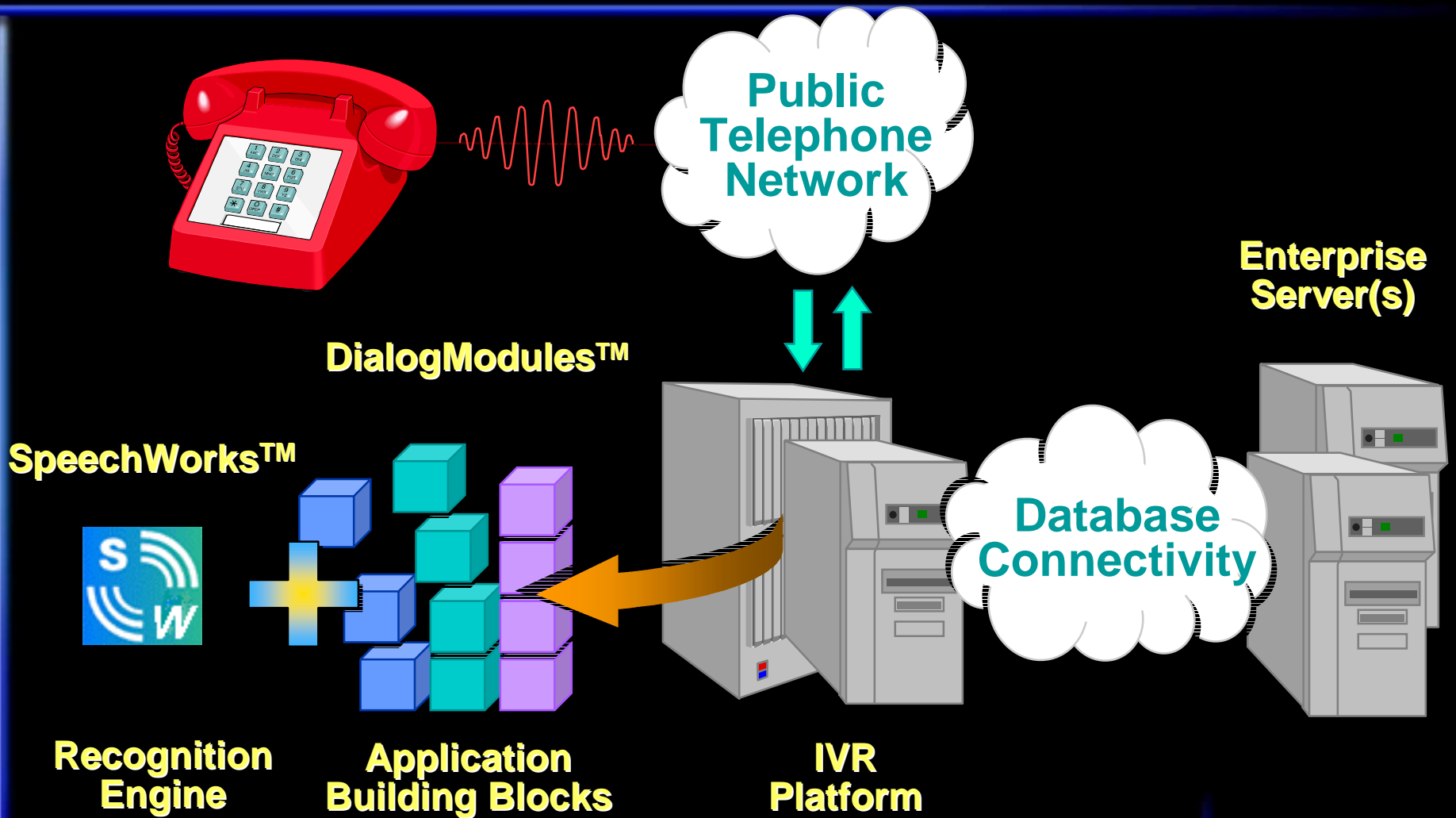


TTS

- TTS becoming more important
 - Dynamic information
 - Voice Portals
 - Large Vocabulary tasks
- Quality now acceptable



Putting It All Together



VoiceXML

- Using Web development approach for speech
 - Application Logic on Web servers
 - Markup language to control speech resources
- Advantages
 - Shares infrastructure with web
 - Platform-independent standard for speech
- New requirements on speech recognizers
 - Dynamic, parallel, cacheable grammars
 - New performance requirements
- SpeechWorks working with CMU to provide open source VoiceXML browser

Building Current Speech Applications

- Technology and Products
- Platforms
- Tools for designing, building, testing, and monitoring
- Application Development Process
 - Lots of attention to User Interface
 - Iterative process of building, testing, tuning
 - Technology important, but no longer prime factor in success

Speech and the Wireless Internet



Wireless Internet So Far

- WAP Phones
 - Small Screens
 - Text-based
 - Low data bandwidth and slow connection speed
 - No simultaneous Voice/Data
 - Tedious data entry
- Wireless PDAs
 - Expensive, unreliable network coverage
 - Small number of devices so far
- DoCoMo

What's Coming

- 2.5G and 3G wireless networks
 - Always on data networks
 - Merged voice/data communications
 - Reasonable bandwidths / costs
- Continued evolution of devices
 - More processing
 - More memory
 - Better displays
- Still tedious data entry

Types of Devices

- Cellphones
 - Mainstream, low-cost handsets
 - Featurephones and Smartphones
- PDAs and Handheld PCs
- Devices built into cars
- Etc.

Next Generation Mobile Devices

- Available Local Computation
- With High Quality Displays
- With Wireless Networking
- Too Small for Keyboard
- Speech + Pointing In
- Speech + Display out
- Local and Distributed Speech Processing



Multi-Modal Interfaces : Dialog-Centric Approach

- Use display and alternate input modalities to improve dialog system
- Display for output - maps, lists, search results, etc.
- Display for dialog state
 - Current state
 - What was understood
 - What is expected next
- Mix of input modalities
- Need to support different devices, situations, user preferences

Multi-Modal Interfaces : Visual-Centric Approach

- Use existing web or other GUI applications
- Augment with speech - speak web links, menu choices, form filling input, etc.
- Augment existing GUI applications with grammars, etc.
- “Dialog” constrained by GUI application
- Need to support different devices, situations, user preferences

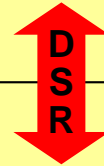
Next Generation Mobile Devices

- Available Local Computation
- With High Quality Displays
- With Wireless Networking
- Too Small for Keyboard
- Speech + Pointing In
- Speech + Display out
- Local and Distributed Speech Processing



Speech Technologies Matrix

	ASR	TTS
<i>Network</i>	<p>Enables</p> <ul style="list-style-type: none">▪ Call Center▪ Voice Portal▪ Multi-modal Access	<p>Enables</p> <ul style="list-style-type: none">▪ Voice Portal▪ E-Mail▪ Knowledge Base
<i>Embedded</i>	<p>Enables</p> <ul style="list-style-type: none">▪ Device command & control▪ Feature extraction/DSR▪ Local data input	<p>Enables</p> <ul style="list-style-type: none">▪ Multi-modal access▪ Device hosted information▪ Streaming data to speech



Future of Speech Technologies



User Interface

- Users will gain experience across interactions
- This experience can be used to form evolving model of user interface
 - Right use of Mixed Initiative, Natural Language
 - Personalization is Key
- Tools will evolve as interface evolves

Speech Technology

- Continued gains on raw technology (30-50% error rate reductions per year)
 - Supports more and more difficult tasks
 - Supports richer User Interfaces
- Wireless and Acoustic robustness increasingly important
 - Network-based approaches
 - Device-based approaches

Conclusion

- More and more demand for mobile access to information and transactions
- Wide range of devices, applications, situations
- Speech input and output is an important mode of interaction
- Development of next generations of wireless infrastructure will have significant impact